

0590  
1012

OIEP

# RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,356

DATE: 11/20/2001

TIME: 11:13:24

Input Set : N:\Crf3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

## SEQUENCE LISTING

### C--> 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Bandman, Olga  
6 Corley, Neil C.  
7 Guegler, Karl J.  
8 Lal, Preeti

### C--> 10 (ii) TITLE OF INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN

12 (iii) NUMBER OF SEQUENCES: 4

14 (iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.  
16 (B) STREET: 3174 Porter Dr.  
17 (C) CITY: Palo Alto  
18 (D) STATE: CA  
19 (E) COUNTRY: USA  
20 (F) ZIP: 94304

22 (v) COMPUTER READABLE FORM:

23 (A) MEDIUM TYPE: Diskette  
24 (B) COMPUTER: IBM Compatible  
25 (C) OPERATING SYSTEM: DOS  
26 (D) SOFTWARE: FastSEQ for Windows Version 2.0

28 (vi) CURRENT APPLICATION DATA:

C--> 29 (A) APPLICATION NUMBER: US/09/866,356

C--> 30 (B) FILING DATE: 29-May-2001

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: 09/055,113  
34 (B) FILING DATE: 1998-04-03

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: Billings, Lucy J.  
40 (B) REGISTRATION NUMBER: 36,749  
41 (C) REFERENCE/DOCKET NUMBER: PF-0501 US

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: 650-855-0555  
45 (B) TELEFAX: 650-845-4166  
46 (C) TELEX:

49 (2) INFORMATION FOR SEQ ID NO: 1:

51 (i) SEQUENCE CHARACTERISTICS:

52 (A) LENGTH: 335 amino acids  
53 (B) TYPE: amino acid  
54 (C) STRANDEDNESS: single  
55 (D) TOPOLOGY: linear

57 (vii) IMMEDIATE SOURCE:

58 (A) LIBRARY: PROSTUT12  
59 (B) CLONE: 1813005

62 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

64 Met Gly Ser Ala Ser Pro Gly Leu Ser Ser Val Ser Pro Ser His Leu  
65 1 5 10 15  
66 Leu Leu Pro Pro Asp Thr Val Ser Arg Thr Gly Leu Glu Lys Ala Ala

ENTERED

## RAW SEQUENCE LISTING

DATE: 11/20/2001

PATENT APPLICATION: US/09/866,356

TIME: 11:13:24

Input Set : N:\Crf3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

```

67          20          25          30
68 Ala Gly Ala Val Gly Leu Glu Arg Arg Asp Trp Ser Pro Ser Pro Pro
69          35          40          45
70 Ala Thr Pro Glu Gln Gly Leu Ser Ala Phe Tyr Leu Ser Tyr Phe Asp
71          50          55          60
72 Met Leu Tyr Pro Glu Asp Ser Ser Trp Ala Ala Lys Ala Pro Gly Ala
73          65          70          75          80
74 Ser Ser Arg Glu Glu Pro Pro Glu Glu Pro Glu Gln Cys Pro Val Ile
75          85          90          95
76 Asp Ser Gln Ala Pro Ala Gly Ser Leu Asp Leu Val Pro Gly Gly Leu
77          100          105          110
78 Thr Leu Glu Glu His Ser Leu Glu Gln Val Gln Ser Met Val Val Gly
79          115          120          125
80 Glu Val Leu Lys Asp Ile Glu Thr Ala Cys Lys Leu Leu Asn Ile Thr
81          130          135          140
82 Ala Asp Pro Met Asp Trp Ser Pro Ser Asn Val Gln Lys Trp Leu Leu
83          145          150          155          160
84 Trp Thr Glu His Gln Tyr Arg Leu Pro Pro Met Gly Lys Ala Phe Gln
85          165          170          175
86 Glu Leu Ala Gly Lys Glu Leu Cys Ala Met Ser Glu Glu Gln Phe Arg
87          180          185          190
88 Gln Arg Ser Pro Leu Gly Gly Asp Val Leu His Ala His Leu Asp Ile
89          195          200          205
90 Trp Lys Ser Ala Ala Trp Met Lys Glu Arg Thr Ser Pro Gly Ala Ile
91          210          215          220
92 His Tyr Cys Ala Ser Thr Ser Glu Glu Ser Trp Thr Asp Ser Glu Val
93          225          230          235          240
94 Asp Ser Ser Cys Ser Gly Gln Pro Ile His Leu Trp Gln Phe Leu Lys
95          245          250          255
96 Glu Leu Leu Leu Lys Pro His Ser Tyr Gly Arg Phe Ile Arg Trp Leu
97          260          265          270
98 Asn Lys Glu Lys Gly Ile Phe Lys Ile Glu Asp Ser Ala Gln Val Ala
99          275          280          285
100 Arg Leu Trp Gly Ile Arg Lys Asn Arg Pro Ala Met Asn Tyr Asp Lys
101          290          295          300
102 Leu Ser Arg Ser Ile Arg Gln Tyr Tyr Lys Lys Gly Ile Ile Arg Lys
103          305          310          315          320
104 Pro Asp Ile Ser Gln Arg Leu Val Tyr Gln Phe Val His Pro Ile
105          325          330          335

```

107 (2) INFORMATION FOR SEQ ID NO: 2:

109 (i) SEQUENCE CHARACTERISTICS:

110 (A) LENGTH: 1905 base pairs

111 (B) TYPE: nucleic acid

112 (C) STRANDEDNESS: single

113 (D) TOPOLOGY: linear

115 (vii) IMMEDIATE SOURCE:

116 (A) LIBRARY: PROSTUT12

117 (B) CLONE: 1813005

119 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

## RAW SEQUENCE LISTING

DATE: 11/20/2001

PATENT APPLICATION: US/09/866,356

TIME: 11:13:24

Input Set : N:\Crf3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

```

121 CTCGAGCCGC GGCTGTCTGA CTTCTCTCCA GCACATTCCT GCACTCTGCC GTGTCCACAC      60
122 TGCCCCACAG ACCCAGTCCT CCAAGCCTGC TGCCAGCTCC CTGCAAGCCC CTCAGGTTGG      120
123 GCCTTGCCAC GGTGCCAGCA GGCAGCCCTG GGCTGGGGGT AGGGGACTCC CTACAGGCAC      180
124 GCAGCCCTGA GACCTCAGAG GGCCACCCCT TGAGGGTGGC CAGGCCCCCA GTGGCCAAAC      240
125 TGAGTGCTGC CTCTGCCACC AGCCCTGCTG GCCCCTGGTT CCGCTGGCCC CCCAGATGCC      300
126 TGGCTGAGAC ACGCCAGTGG CCTCAGCTGC CCACACCTCT TCCCGGCCCC TGAAGTTGGC      360
127 ACTGCAGCAG ACAGCTCCCT GGGCACCAGG CAGCTAACAG ACACAGCCGC CAGCCCCAAC      420
128 AGCAGCGGCA TGGGCAGCGC CAGCCCGGGT CTGAGCAGCG TATCCCCCAG CCACCTCTCTG      480
129 CTGCCCCCCG ACACGGTGTG GCGGACAGGC TTGGAGAAGG CGGCAGCGGG GGCAGTGGGT      540
130 CTCGAGAGAC GGGACTGGAG TCCCAGTCCA CCCGCCACGC CCGAGCAGGG CCTGTCCGCC      600
131 TTCTACCTCT CTTACTTTGA CATGCTGTAC CCTGAGGACA GCAGCTGGGC AGCCAAGGCC      660
132 CCTGGGGCCA GCAGTCGGGA GGAGCCACCT GAGGAGCCTG AGCAGTGCCC GGTCAATTGAC      720
133 AGCCAAGCCC CAGCGGGCAG CCTGGACTTG GTGCCCGGCG GGCTGACCTT GGAGGAGCAC      780
134 TCGCTGGAGC AGGTGCAGTC CATGGTGGTG GCGAAGTGC TCAAGGACAT CGAGACGGCC      840
135 TGCAAGCTGC TCAACATCAC CGCAGATCCC ATGGACTGGA GCCCCAGCAA TGTGCAGAAG      900
136 TGGCTCCTGT GGACAGAGCA CCAATACCGG CTGCCCCCA TGGGCAAGGC CTTCCAGGAG      960
137 CTGGCGGGCA AGGAGCTGTG CGCATGTG GAGGAGCAGT TCCGCCAGCG CTCGCCCTG      1020
138 GGTGGGGATG TGCTGCACGC CCACCTGGAC ATCTGGAAGT CAGCGGCCTG GATGAAAGAG      1080
139 CGACTTCAC CTGGGGCGAT TCACTACTGT GCCTCGACCA GTGAGGAGAG CTGGACCGAC      1140
140 AGCGAGGTGG ACTCATCATG CTCCGGGCAG CCCATCCACC TGTGGCAGTT CCTCAAGGAG      1200
141 TTGCTACTCA AGCCCCACAG CTATGGCCGC TTCATTAGGT GGCTCAACAA GGAGAAGGGC      1260
142 ATCTTCAAAA TTGAGGACTC AGCCCAGGTG GCCCAGCTGT GGGGCATCCG CAAGAACCGT      1320
143 CCCGCCATGA ACTACGACAA GCTGAGCCGC TCCATCCGCC AGTATTACAA GAAGGGCATC      1380
144 ATCCGGAAGC CAGACATCTC CCAGCGCCTC GTCTACCACT TCGTGACCCC CATCTGAGTG      1440
145 CCTGGCCAGG GGCCTGAAAC CCGCCCTCAG GGGCCTCTCT CCTGCCTGCC CTGCCTCAGC      1500
146 CAGGCCCTGA GATGGGGGAA AACGGGCAGT CTGCTCTGCT GCTCTGACCT TCCAGAGCCC      1560
147 AAGGTCAGGG AGGGGCAACC AACTGCCCA GGGGATATG GGTCCTCTGG GCCTTCGGG      1620
148 ACCCTGGGGC AGGGGTGCTT CCTCCTCAGG CCCAGCTGCT CCCCTGGAGG ACAGAGGGAG      1680
149 ACAGGGCTGC TCCCCAACAC CTGCCTCTGA CCCCAGCATT TCCAGAGCAG AGCCTACAGA      1740
150 AGGGCAGTGA CTCGACAAAG GCCACAGGCA GTCCAGGCCT CTCTCTGCTC CATCCCCCTG      1800
151 CCTCCCATTG TGCAACACAC CTGGCATGGT GCAGGGAGAC ATCTGCACCC CTGAGTTGGG      1860
152 CAGCCAGGAG TGCCCCCGGG AATGGATAAT AAAGATACTA GGCGC      1905

```

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

(A) LIBRARY: GenBank

(B) CLONE: 157196

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

156 Thr Asn Ala Ser Asn Gly Gly Thr Ala Thr Val Lys Arg Pro Asn Gly
157 1      5      10      15
158 Gly Arg Thr Gly Gly Gly Gly Ser His Ile His Leu Trp Gln Phe Leu
159 20     25     30
160 Lys Glu Leu Leu Ala Ser Pro Gln Val Asn Gly Thr Ala Ile Arg Trp
161 35     40     45
162 Ile Asp Arg Ser Lys Gly Ile Phe Lys Ile Glu Asp Ser Val Arg Val

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,356

DATE: 11/20/2001

TIME: 11:13:24

Input Set : N:\Crif3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

```

175      50      55      60
176 Ala Lys Leu Trp Gly Arg Arg Lys Asn Arg Pro Ala Met Asn Tyr Asp
177 65      70      75      80
178 Lys Leu Ser Arg Ser Ile Arg Gln Tyr Tyr Lys Lys Gly Ile Met Lys
179      85      90      95
180 Lys Thr Glu Arg Ser Gln Arg Leu Val Tyr Gln Phe Cys His Pro Tyr
181      100      105      110
182 Ser Gln

```

185 (2) INFORMATION FOR SEQ ID NO: 4:

187 (i) SEQUENCE CHARACTERISTICS:

188 (A) LENGTH: 452 amino acids

189 (B) TYPE: amino acid

190 (C) STRANDEDNESS: single

191 (D) TOPOLOGY: linear

193 (vii) IMMEDIATE SOURCE:

194 (A) LIBRARY: GenBank

195 (B) CLONE: 511283

197 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

199 Met Ser Glu Thr Pro Ala Gln Cys Ser Ile Lys Gln Glu Arg Ile Ser
200 1      5      10      15
201 Tyr Thr Pro Pro Glu Ser Pro Val Pro Ser Tyr Ala Ser Ser Thr Pro
202      20      25      30
203 Leu His Val Pro Val Pro Arg Ala Leu Arg Met Glu Glu Asp Ser Ile
204      35      40      45
205 Arg Leu Pro Ala His Leu Arg Leu Gln Pro Ile Tyr Trp Ser Arg Asp
206      50      55      60
207 Asp Val Ala Gln Trp Leu Lys Trp Ala Glu Asn Glu Phe Ser Leu Arg
208 65      70      75      80
209 Pro Ile Asp Ser Asn Thr Phe Glu Met Asn Gly Lys Ala Leu Leu Leu
210      85      90      95
211 Leu Thr Lys Glu Asp Phe Arg Tyr Arg Ser Pro His Ser Gly Asp Val
212      100      105      110
213 Leu Tyr Glu Leu Leu Gln His Ile Leu Lys Gln Arg Lys Pro Arg Ile
214      115      120      125
215 Leu Phe Ser Pro Phe Phe His Pro Gly Asn Ser Ile His Thr Gln Pro
216      130      135      140
217 Glu Val Ile Leu His Gln Asn His Glu Glu Asp Asn Cys Val Gln Arg
218 145      150      155      160
219 Thr Pro Arg Pro Ser Val Asp Asn Val His His Asn Pro Pro Thr Ile
220      165      170      175
221 Glu Leu Leu His Arg Ser Arg Ser Pro Ile Thr Thr Asn His Arg Pro
222      180      185      190
223 Ser Pro Asp Pro Glu Gln Arg Pro Leu Arg Ser Pro Leu Asp Asn Met
224      195      200      205
225 Ile Arg Arg Leu Ser Pro Ala Glu Arg Ala Gln Gly Pro Arg Pro His
226      210      215      220
227 Gln Glu Asn Asn His Gln Glu Ser Tyr Pro Leu Ser Val Ser Pro Met
228 225      230      235      240
229 Glu Asn Asn His Cys Pro Ala Ser Ser Glu Ser His Pro Lys Pro Ser

```

## RAW SEQUENCE LISTING

DATE: 11/20/2001

PATENT APPLICATION: US/09/866,356

TIME: 11:13:24

Input Set : N:\Crf3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

230		245		250		255
231	Ser Pro Arg Gln Glu Ser Thr Arg Val Ile Gln Leu Met Pro Ser Pro					
232		260		265		270
233	Ile Met His Pro Leu Ile Leu Asn Pro Arg His Ser Val Asp Phe Lys					
234		275		280		285
235	Gln Ser Arg Leu Ser Glu Asp Gly Leu His Arg Glu Gly Lys Pro Ile					
236		290		295		300
237	Asn Leu Ser His Arg Glu Asp Leu Ala Tyr Met Asn His Ile Met Val					
238	305		310		315	320
239	Ser Val Ser Pro Pro Glu Glu His Ala Met Pro Ile Gly Arg Ile Ala					
240		325		330		335
241	Asp Cys Arg Leu Leu Trp Asp Tyr Val Tyr Gln Leu Leu Ser Asp Ser					
242		340		345		350
243	Arg Tyr Glu Asn Phe Ile Arg Trp Glu Asp Lys Glu Ser Lys Ile Phe					
244		355		360		365
245	Arg Ile Val Asp Pro Asn Gly Leu Ala Arg Leu Trp Gly Asn His Lys					
246		370		375		380
247	Asn Arg Thr Asn Met Thr Tyr Glu Lys Met Ser Arg Ala Leu Arg His					
248	385		390		395	400
249	Tyr Tyr Lys Leu Asn Ile Ile Arg Lys Glu Pro Gly Gln Arg Leu Leu					
250		405		410		415
251	Phe Arg Phe Met Lys Thr Pro Asp Glu Ile Met Ser Gly Arg Thr Asp					
252		420		425		430
253	Arg Leu Glu His Leu Glu Ser Gln Glu Leu Asp Glu Gln Ile Tyr Gln					
254		435		440		445
255	Glu Asp Glu Cys					
256		450				

## VERIFICATION SUMMARY

DATE: 11/20/2001

PATENT APPLICATION: US/09/866,356

TIME: 11:13:25

Input Set : N:\Crf3\RULE60\09866356.txt

Output Set: N:\CRF3\11202001\I866356.raw

L:3 M:220 C: Keyword misspelled or invalid format, [(1) GENERAL INFORMATION:]  
L:10 M:220 C: Keyword misspelled or invalid format, [(ii) TITLE OF INVENTION:]  
L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]